

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Papers in Natural Resources

Natural Resources, School of

10-2012

A geospatial modeling framework for assessing biofuels-related land-use and land-cover change

Ruopu Li

University of Nebraska-Lincoln, ruopulee@yahoo.com

Qingfeng (Gene) Guan

University of Nebraska - Lincoln, qguan2@unl.edu

James W. Merchant

University of Nebraska-Lincoln, jmerchant1@unl.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/natrespapers>



Part of the [Natural Resources and Conservation Commons](#)

Li, Ruopu; Guan, Qingfeng (Gene); and Merchant, James W., "A geospatial modeling framework for assessing biofuels-related land-use and land-cover change" (2012). *Papers in Natural Resources*. 332. <https://digitalcommons.unl.edu/natrespapers/332>

This Article is brought to you for free and open access by the Natural Resources, School of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Papers in Natural Resources by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

A geospatial modeling framework for assessing biofuels-related land-use and land-cover change

Ruopu Li, Qingfeng Guan, James Merchant

Center for Advanced Land Management Information Technologies, School of Natural Resources,
University of Nebraska – Lincoln, 3310 Holdrege Street, Lincoln, NE 68583-0973, USA

Corresponding author — Q. Guan, email qguan2@unl.edu

Abstract

With the increasing biofuel demands in recent years, the cultivated lands for biofuel crops, i.e., corn and soybeans, the major sources of ethanol and biodiesel, have been greatly expanding in the northern Great Plains states of the United States. Simulating the spatio-temporal dynamics of biofuel croplands can provide critical information required for assessing the impacts of land-use change on wildlife conservation and water quality. But, yearly agricultural practices such as crop rotations often complicate the spatially explicit modeling of specific crops' expansion. Our research focused on developing a geospatial modeling framework that is able to distinguish long-term, regional changes in croplands from short-term, local fluctuations (such as rotations), using geographic information systems (GIS) and the land transformation model (LTM). The USDA Cropland Data Layers (CDLs) of North Dakota for 1999, 2000, 2004, 2005, 2010, and 2011 were spatially and temporally aggregated to generate a series of biofuel cropland maps. The historical cropland data for 1999/2000 and 2004/2005, together with a collection of environmental factors (i.e., topography, soil fertility, and climate), were used to calibrate the neural network embedded in the LTM. Validation analysis was then conducted by simulating the biofuel cropland change during the period of 2004/2005–2010/2011 using the calibrated LTM and comparing the simulation result with the observed change for the same time period, resulting in 6.3% allocation disagreement (0% quantity disagreement) and 27.4% figure of merit. Future forecast for 2020 showed that biofuel croplands would be expanding northwestward from southeastern North Dakota.

Keywords: Land-Use and Land-Cover Change (LULCC), Land Transformation Model (LTM), Corn, Soybeans, Biofuels, Biofuel crops, North Dakota

1. Introduction

In the last decade, the land devoted to growing corn and soybeans in the northern Great Plains states (including Iowa, Nebraska, Minnesota, South Dakota, and North Dakota) of the United States has greatly expanded (Table 1). An important driver of this expansion is the increasing demand for biofuels (Wallander et al., 2011; Carriquiry, 2007; Secchi and Babcock, 2007). Demands for corn, used to produce bioethanol, and soybeans, used to produce biodiesel, are expected to be strong in the foreseeable future (Woodard, 2009). The expansion of corn and soybeans cultivation can cause a spectrum of negative environmental and ecological consequences (Kennedy, 2007; de Oliveira et al., 2005). For example, increases in corn and soybeans croplands may lead to pressures to remove land from the Conservation Reserve Program (CRP), drain wetlands, and open lands that are currently not cultivated, and hence cause loss of critical wildlife habitat (Brooke et al., 2009).

In addition, the expansion of corn and soybeans croplands can affect the quality of both surface and groundwater because biofuel

crop cultivation generally requires much higher inputs of fertilizer and other farm chemicals that can be flushed into water bodies or leach into groundwater (Thomas et al., 2009). The deterioration of water quality accompanying such land-use conversion is a major threat to both human health and ecosystems.

Therefore, better understanding of the spatio-temporal dynamics of biofuel croplands, its drivers and consequences is critical to the development of effective environmental management strategies. An important component of such work is to develop viable models to simulate biofuels-related land-use and land-cover changes (LULCC).

A number of geospatial models have been developed to simulate patterns and processes of LULCC (Pontius et al., 2008). Models such as SLEUTH (Clarke et al., 1997), the land transformation model (LTM) (Pijanowski et al., 2002, 2005), and CLUE/CLUE-S (Verburg et al., 1999, 2002) have been widely applied to forecast loss of agricultural land (Fan et al., 2007) and urban growth (Pijanowski et al., 2002; Clarke et al., 1997), but have rarely been used for forecasting the spatio-temporal changes in specific cultivated croplands. Even though some efforts were made to model

Table 1. Expansion of corn and soybeans land between 1988 and 2008 in five states (unit: thousand hectares).

State	Crop/year			
	Corn/1988	Corn/2008	Soybeans/1998	Soybeans/2008
Iowa	4573.1	5382.5	3298.3	3945.8
Minnesota	2306.8	3116.2	1983	2853.1
Nebraska	2792.4	3561.4	971.3	1983
North Dakota	323.8	1032	303.5	1537.9
South Dakota	1274.8	1922.3	712.3	1659.3

Data Source: http://www.nass.usda.gov/Data_and_Statistics/

changes in specific crops, they have often been founded based on data compiled for highly aggregated spatial units such as counties, statistical districts or countries. For example, de la Torre Ugarte and Ray (2000) used the Policy Analysis System (POLYSYS), a complex economic demand-supply model, to estimate the national distribution of U.S. bioenergy crops at the Agricultural Statistics District (ASD) level. Smeets et al. (2006) used the Quickscan model to forecast bioenergy crop production in 2050 at the country level with a global coverage. Modeling based on such coarse, aggregated units often masks local variations and therefore has low utility for managing the environmental and ecological consequences of LULCC (Verburg et al., 1999).

In recent years, a few attempts have been made to model biofuels-related cropland change using location-specific approaches. Tuck et al. (2006) mapped the potential distribution of bioenergy crops in Europe based on a set of simple rules defining suitable climate conditions and elevation. Hallmann and Verburg (2011) used a spatially explicit model to forecast changes in biofuel croplands in Europe. Evans et al. (2010) assessed the landscape suitability for growing biofuel feedstocks in the United States based on two species distribution models: suitability maximum entropy (Maxent) and support vector machines (SVM). These studies have provided constructive insights into the spatially explicit modeling of biofuel croplands. Nevertheless, two important issues remain to be resolved: (1) long-term, regional trends, which are usually the focus of modeling, are often mixed with short-term, local fluctuations arising from management practices such as crop rotation; (2) the validation of modeling results should also be conducted in a spatially explicit manner, i.e., statistically comparing the modeled distribution of change in cropland with the historical maps of actual change, in order to examine the accuracy of the model before it is used to generate future forecasts.

Therefore, the principal objective of this research was to develop a spatially explicit modeling framework that is capable of operating at multiple spatial and temporal scales in order to distinguish long-term, regional trends from short-term, local fluctuations. The model was used to simulate the recent history of corn and soybeans cropland expansion in North Dakota, and the result was assessed by comparing it with actual historical data. Future scenarios of cropland change were also generated. Since issues such as competing use of corn and soybeans for fuels and food/livestock crops are not the focus of this research, we did not distinguish the croplands according to their usage in the modeling work. Thus, the modeled corn and soybeans can be used for both fuels and food purposes.

2. Spatial and temporal scales in LULCC modeling

Before introducing the methods used in our modeling framework, it is important to discuss the effects of spatial and tempo-

ral scales in LULCC modeling because that is the focus of this research. Scale is an inherent attribute of geographic phenomena (Verburg et al., 1999; Cao and Lam, 1997). In studies of LULCC, both spatial and temporal scales must be considered. Moreover, one must account for both extent (i.e., the entire study area or modeling time period) and resolution (i.e., the smallest mapping unit or time interval represented in the dataset). Figure 1 illustrates a study area with different spatial resolutions and extents. As the spatial resolution increases¹ (from 4×4 units to 1×1 units in Figure 1), one can portray features with greater detail; as the spatial extent increases² (ninefold in Figure 1), the study area becomes larger. In LULCC studies, spatial extent and resolution are often related. For example, at smaller spatial extents it is common to encode data at finer (i.e., higher) spatial resolution so that details can be discerned. At larger extents, the spatial resolution may be coarsened (i.e., lowered) to reveal more general LULC patterns. Similarly, the extent and resolution of temporal scale are also important in the analysis of changes. From a temporal perspective, the extent is the length of the entire time period analyzed, while the resolution is expressed as the smallest time interval utilized in analysis (e.g., day, month, season, or year). In general, LULCC observations made over short-time intervals (i.e., fine temporal resolution) are required for intra-annual analysis and modeling, while data with coarser temporal resolution (e.g., annual or longer) are often acceptable for long-term (e.g., inter-annual and decadal) studies.

Croplands are complex and dynamic systems that can be represented at different spatial and temporal scales. Changes in scale can largely affect the observation of LULCC and related spatial patterns (Goodchild and Dale, 1997; Turner, 1990), and hence affect the modeling of LULCC. Figure 2 shows how observations of croplands vary as spatial and temporal extents change. Studies of crop physiological changes usually are conducted with data with limited spatial and temporal extents (e.g., within several square meters and over a few seasons) and fine resolutions (e.g., individual crop and daily observation). Moderate spatial and temporal scales are better suited for studies of planting rotational patterns. Coarse resolutions are usually best for research focused on long-term agricultural LULCC, and crop rotations should be masked at these scales.

LULCC models are therefore inherently scale-dependent (Bian, 1997). To model regional agricultural LULCC, it is necessary to minimize the effects of crop rotations and other short-term, local LULCC driven by fluctuating crop markets and agricultural policies. Such short-term, local changes can introduce substantial year-to-year "noise", making it difficult to analyze and model long-term LULCC patterns at a regional level. Our modeling framework includes a spatial and temporal aggregation process of the LULC data in order to smooth the noise caused by those short-term, local fluctuations in croplands.

1. Increasing spatial resolution means using smaller mapping units.

2. Increasing spatial extent means expanding the coverage of the map.

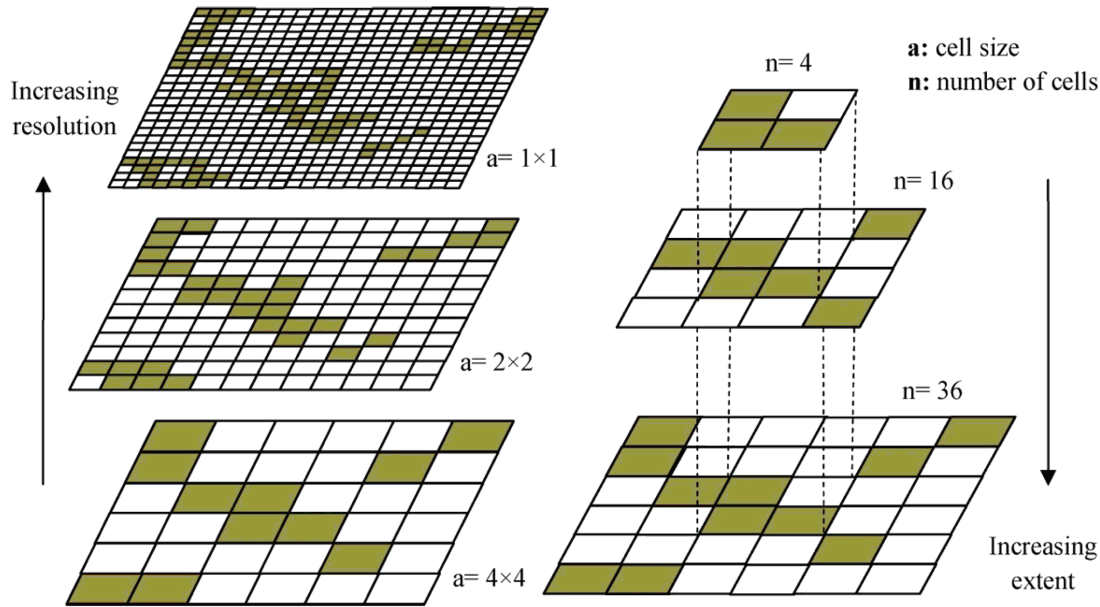


Figure 1. Changes in two components of spatial scale: resolution (left) and extent (right).

3. Methodology

In this research, a grid-based geospatial modeling approach was developed to simulate and forecast the land-use changes of two important biofuel crops: corn and soybeans. In the northern Great Plains, corn and soybeans are commonly grown in rotation. Since the pattern of corn-soybeans rotation at a specific location is extremely difficult to predict, our model treats these two crops as a single land-use class, i.e., “biofuel crop”.³ Our approach was designed to capture general patterns and trends, and generate location-specific results, while reducing the effects of short-term, local fluctuations (e.g., crop rotations) associated with croplands. The model was then evaluated by employing it to simulate the recent history of corn and soybeans expansion in North Dakota.

The modeling framework includes two key modules: the quantity module and the spatial allocation module (Figure 3). The quantity module is used to determine and forecast the total amount of changes in corn/soybeans croplands (i.e., the number of land cells of other land-use types to be transformed into corn/soybeans cropland) during a particular time period. The spatial allocation module is then used to spatially distribute these changes (i.e., to determine which specific cells in the grid to transform). This approach was based on a common assumption in LULCC modeling that spatially explicit geographic processes can be constrained by less spatially precise economic or policy-making processes (Lambin et al., 2000; Verburg et al., 1999).

Forecasting the future changes in croplands is complicated by several issues: (1) corn and soybeans serve multiple functions (e.g., as biofuels, food, and other commodities) and compete with other crops; (2) biofuel demands can be affected by factors both inside and outside the region; and (3) crop yields will vary in response to both marketing conditions, and weather and climate events. The quantity module, as implemented in this study, projects future corn/soybeans demands using statistical extrapolation of historical trends in crop area. This approach, while simple, is computationally straightforward and is believed to capture important components of the factors outlined above without the need to independently model each.

3. There are other biofuel crops, such as canola, sugar beets and sunflowers. In this research, we focused on only two main biofuel crops in the northern Great Plains of U.S. – corn and soybeans.

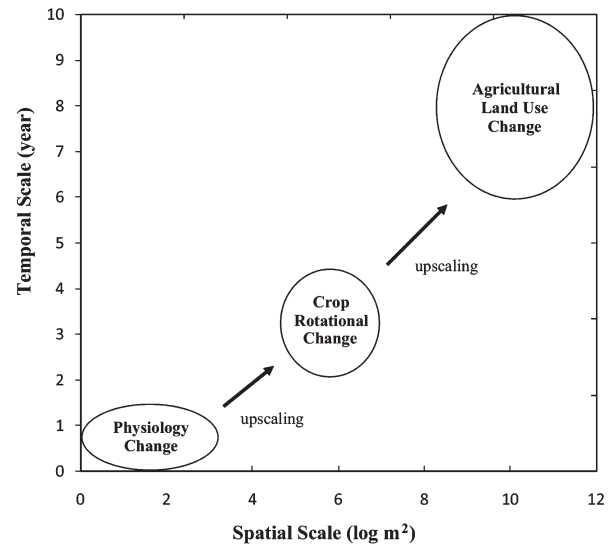


Figure 2. Agricultural land use change processes affected by spatial and temporal scales.

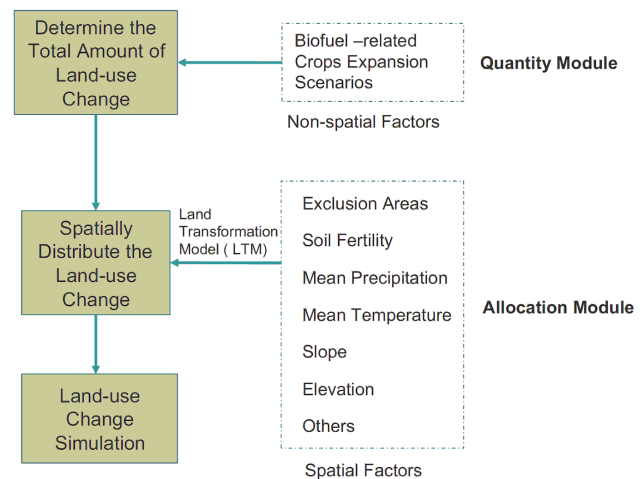


Figure 3. Framework for modeling cropland change.

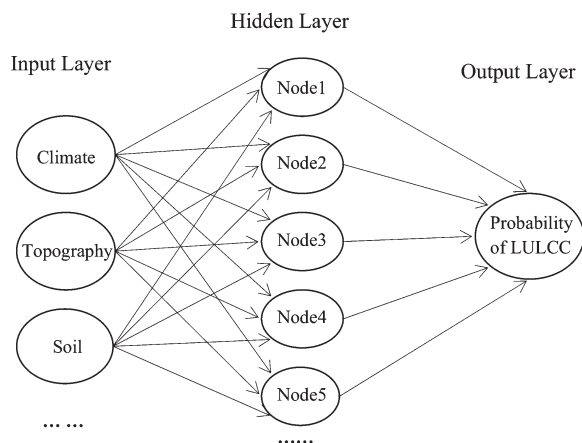


Figure 4. A simple 3-layer artificial neural network.

The spatial allocation module is based on an existing LULCC model, the land transformation model (LTM). The LTM is a grid-based spatially explicit, well-tested and freely available model that integrates environmental and socio-economic drivers with historical land-use datasets to simulate LULCC (Pijanowski et al., 2002). The core of the LTM is an Artificial Neural Network (ANN), which uses a machine learning approach for modeling complex systems and dynamics (Haupt et al., 2009). The ANN consists of an input layer comprised of a set of nodes representing driving factors, an output layer consisting of only one node that represents the suitability for a certain land-use type (e.g., urban land in an urban growth study, or biofuel cropland in this study), and one or more hidden layers in between (Figure 4). The nodes within adjacent layers are connected through active transfer functions (ATFs).

Most previous applications of the LTM have focused on simulating urban growth (Pijanowski et al., 2002, 2005; Tang et al., 2005), but the flexibility of the ANN embedded in the LTM allows for the simulation of other types of LULCC. Technically, any factor can be used as an input to an ANN, and the output can be any variable that is of interest in a study. The selection of input factors and output variable(s) largely depends on the purpose of the simulation and the mechanism of the phenomenon. Moreover, an ANN is capable of handling complex non-linear relationships between the input factors and the output variable(s), and of acquiring knowledge from incomplete, redundant, and noisy datasets without pre-defined rules, both of which are characteristics common in models of LULCC (Gosav and Praisler, 2008; Kajita et al., 2005; Mas et al., 2004; Pijanowski et al., 2002; Hilbert and Ostendorf, 2001).

Through a learning (also termed training or calibration) process using historical datasets, the LTM ANN adjusts the weights of ATFs to establish functional relationships between the driving factors and land-use conversion. In other words, the ANN “learns” by acquiring knowledge based on the past history of land-use change. Once trained, the ANN can be used to simulate land-use change either retroactively, by attempting to replicate past observed changes, or to forecast future changes.

In this research, the LTM is employed to model corn/soybeans cropland changes. The probability of transforming a cell from other land-use types to corn/soybeans cropland is set to be the output of the ANN and a flexible set of factors (e.g., slope and soil organic matter) that may affect cropland expansion are selected as the inputs. The model essentially generates a suitability map for biofuel cropland, and then selects the cells exhibiting the highest suitability to convert. Also, the LTM allows users to specify the number of cells to convert during the modeling period. By setting the numbers of cells converting to corn/soybeans cropland in the calibration and history simulation (i.e., validation) processes as the numbers of ob-

served changed cells derived from the historical data, we eliminated the quantity disagreement between the simulated and observed (or historical) LULCC, which is one of the two map comparison metrics (the other one is allocation disagreement) suggested by Pontius et al. (2008). In the forecasting process, the number of cells to convert is set according to the future projection of corn/soybeans demand generated by the quantity module.

Pijanowski et al. (2002) described six steps of applying the LTM: (1) mapping historical land use; (2) identifying driving factors; (3) preprocessing the raster layers for both land use and driving factors; (4) testing the model with all inputs; (5) calibrating and validating the model; and (6) identifying transitional cells to create possible scenarios of future land use. In this study, the modeling performance was evaluated by examining the disagreement between the simulated and observed LULCC, quantified using two statistics, namely allocation disagreement (quantity disagreement was eliminated as explained above) and figure of merit (for details, see Pontius and Millones, 2011; Pontius et al., 2008). The Percent Correct Metric (PCM) and Kappa, statistical metrics commonly used in evaluating LULCC modeling quality and also provided by the LTM, were excluded from the modeling performance assessment since recent research found them redundant, misleading and flawed for map comparison and accuracy assessment (Pontius and Millones, 2011).

As mentioned in Section 2, our modeling framework focuses on the long-term, regional trends of biofuels-related LULCC. To minimize the effect of short-term, local fluctuations such as crop rotations, the following procedure was employed to prepare data for the LTM (Figure 5).

- (1) Reclassify the original fine-resolution (e.g., 30 m \times 30 m) land-use data into a binary representation, i.e., the target cropland (e.g., corn and soybeans) as value 1, and other land-use types as value 0.
- (2) Aggregate the binary data into coarse-resolution (e.g., 1500 m \times 1500 m) cells and assign each cell a value representing the areal percentage of the target cropland.
- (3) Average the areal percentages of target cropland from multiple land-use data for consecutive years, to generate new land-use data with a multi-year temporal resolution (Figure 5 shows 2-year temporal averaging).
- (4) Finally, reclassify the averaged data into a binary representation, i.e., an averaged percentage greater than or equal to a pre-set threshold reclassified to value 1, otherwise 0.

The selection of spatial and temporal resolution is critically important as the ideal resolution varies according to the application (e.g., target cropland, data availability, and spatio-temporal ranges of crop rotations and market fluctuations). In the case study of corn and soybeans croplands in North Dakota, the spatial resolution was set to approximate the size of a U.S. Public Land Survey section (1500 m \times 1500 m), which minimized the LULCC variability while maintaining a credible level of spatial explicitness. The temporal resolution was set to 2 years in order to reduce the impacts of short-term (i.e., inter-annual) fluctuations in LULCC stemming from annual crop rotations as well as climatic anomalies and volatile crop market conditions.

Also, the threshold value used to reclassify areal percentage maps to binary maps has to be set with caution. Once reclassified, a cell is considered as a uniform land use, and the variation within the cell is lost. Applying a lower threshold produces more biofuel cropland cells and hence provides a larger land-use change contrast for the model, which may lead to a better goodness-of-fit for the simulation. On the other hand, a lower threshold exaggerates the quantity of biofuel croplands (e.g., a cell with just a tiny fraction for corn/soybean cultivation will be classified into

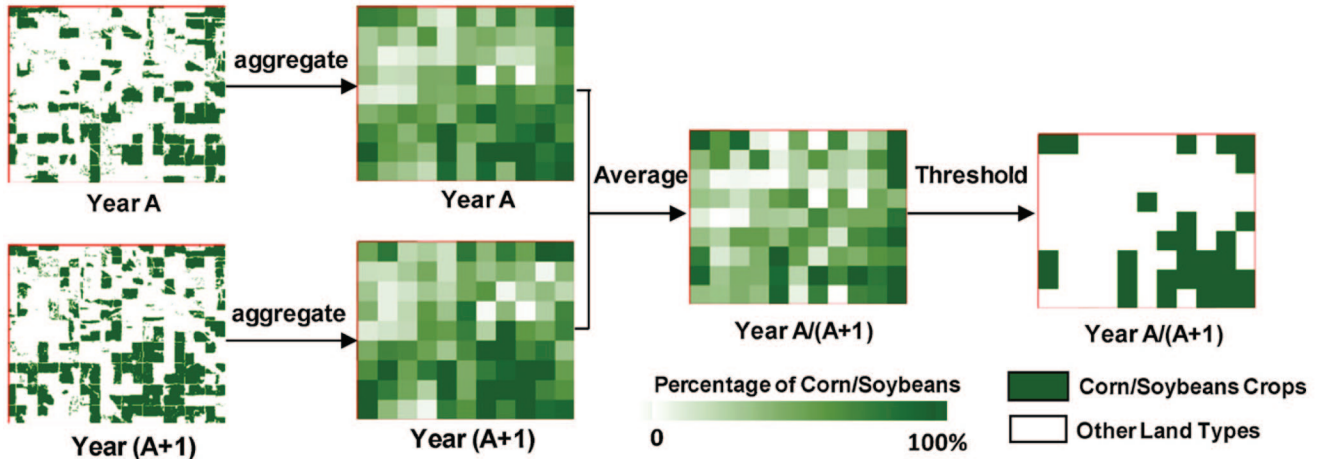


Figure 5. Spatio-temporal aggregations

biofuel cropland), which makes the simulation less realistic. An optimal threshold should provide a good modeling fit while representing a reasonable fraction of the target land-use class (e.g., biofuel crops) within each cell. A similar approach has been adopted for land-use representation by the CLUE-S model, resulting in one dominant (>50%) land type occupying pixels of land-use map (Verburg et al., 2002). In this study, a threshold of 40% was chosen for the study of North Dakota in consideration of the trade-off between the goodness of modeling fit and the representative fraction of biofuel crops within each cell. More discussion on this issue will be given in Section 6.

4. Application of the model in North Dakota

4.1. Study area

North Dakota was selected as the study area because it is representative of the northern Great Plains states, a region that has been experiencing substantial changes in land use thought to be driven in part by increasing demand for biofuels (Table 1). North Dakota has a continental climate typified by cold winters and hot summers; however, during the past century average temperatures in North Dakota have increased by up to 3 °C (U.S. Global Change Research Program, 2000), among the highest in the northern Great Plains. The state is the leading producer

of wheat, barley, sunflowers and dry edible beans in the U.S. However, since the late 1990s, cropland change in North Dakota has been characterized by the rapid expansion of corn and soybeans croplands (Schnitkey, 2010). Corn and soybeans have generally either displaced other crops (such as wheat and sunflowers) or been planted on lands formerly in the Conservation Reserve Program (CRP). In 1997, the top three agricultural commodities were wheat, cattle and sunflower, accounting for 39.3%, 12% and 8.3% of the state total farm receipts, respectively. By 2008, however, the three most important farm commodities changed to wheat, soybeans and corn at 33%, 14.4% and 14.3%, respectively (Economic Research Service, 2009).

4.2. Data preprocessing

A time series (1999–2011) of land-use data for North Dakota were obtained from the U.S. Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) Cropland Data Layers (CDLs) — <http://www.nass.usda.gov/research/Cropland/SARS1a.htm>. Most of 1999–2011 CDLs are available at 30 m resolution, and 2006–2009 CDLs are available at 56 m resolution. As noted above, corn and soybeans were combined into a single land-use type in this study. Following the procedure shown in Figure 5, the CDLs for the years 1999, 2000, 2004, 2005, 2010 and 2011 were first reclassified into binary data (corn/soybeans as 1, and oth-

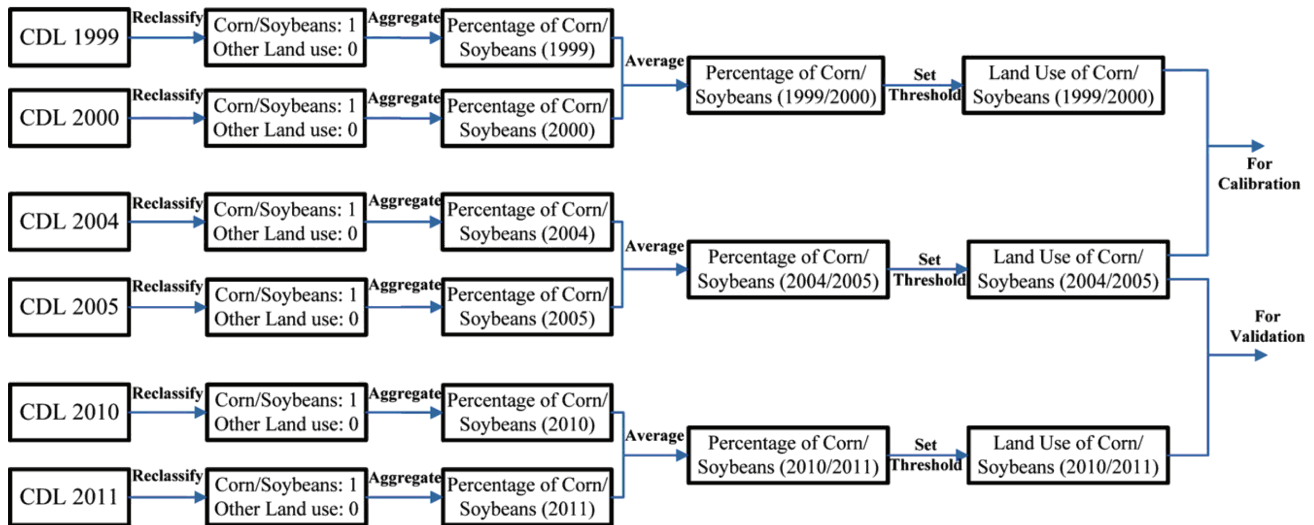


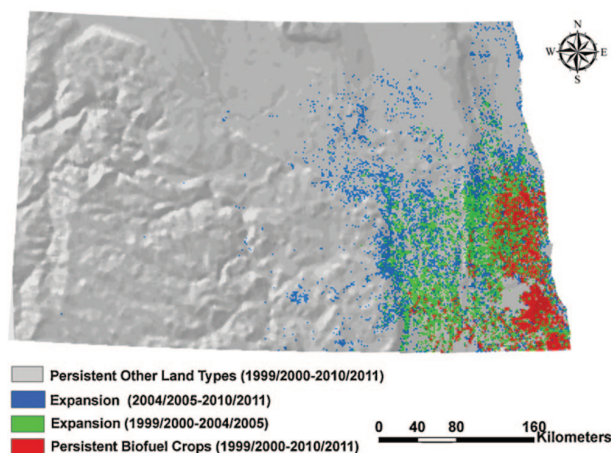
Figure 6. Procedure to prepare land-use maps for modeling.

Table 2. Factors used to predict cropland change.

Factor	Relationship to cropland change	Data source
Elevation and slope	Topography influences water availability, and physical and chemical properties of soil (Kravchenko and Bullock, 2000) which can affect crop yields	USGS National Elevation Dataset
Soil organic matter and CEC	Organic matter can release plant nutrients, including nitrogen and phosphorus as it is broken down in the soil. CEC can affect the soil's capacity to hold nutrients releasable for plant growth (Griffin, 2004)	USDA NRCS STATSGO Database
Mean precipitation and mean temperature	Precipitation is generally related to the spatial distribution of soil moisture, which is important for agricultural cultivation. Annual mean temperature can affect crops' temperature requirements for growth	PRISM Climate Group http://www.prism.oregonstate.edu

ers as 0); then the binary data (at 30 m spatial resolution) were aggregated in ArcGIS (a commonly used GIS software package) to generate 1500-m resolution grids of corn/soybeans areal percentages (Figure 6). The percentages were then averaged between 1999 and 2000, 2004 and 2005, 2010 and 2011 to produce three maps for 1999/2000, 2004/2005 and 2010/2011. Cells in each 2-year map were subsequently reclassified using the following rules: all cells that contained at least 40% corn and soybeans were reclassified to corn/soybeans cells (value = 1), while other cells were given the value zero. This procedure resulted in the biofuel cropland map for 1999/2000, 2004/2005 and 2010/2011 (Figure 7).

Six environmental variables were chosen as the driving factors for modeling the biofuel cropland changes in North Dakota (Table 2): terrain elevation, terrain slope, soil organic matter, cation exchange capacity (CEC) of the soil, 30-year mean precipitation (1971–2000), and 30-year mean temperature (1971–2000). All are important to establishing the suitability of land for growing crops (Bowen and Hollinger, 2002; Kravchenko and Bullock, 2000). Elevation and slope data were derived from the USGS National Elevation Dataset, and resampled into 1500-m resolution grids. Soil organic matter and CEC were extracted from the USDA STATSGO database using the Soil Data Viewer — <http://soils.usda.gov/sdv/>. The shapefiles (a GIS data format used in ArcGIS) of soil organic matter and CEC were then converted into 1500-m resolution grids. The mean precipitation and mean temperature were extracted from Parameter-elevation Regressions on the Independent Slopes Model (PRISM) climate mapping system — <http://www.prism.oregonstate.edu/> — and converted into 1500-m resolution grids. Exclusionary zones (including public lands, wetlands, urban areas and water bodies) were extracted from the Land Ownership for the Western United States — <http://sagemap.wr.usgs.gov/> — and 2000 CDL.

**Figure 7.** Biofuel cropland (corn/soybeans) maps for 1999/2000, 2004/2005 and 2010/2011 (data were derived based on CDLs for North Dakota).**Table 3.** Two-by-two contingency table showing the proportion of grid cells in a map of reality versus a map of calibration scenario (1999/2000–2004/2005).

Model	Reality		
	Change	Persistence	Total
Change	5.03%	2.78%	7.81%
Persistence	2.78%	89.41%	92.19%
Total	7.81%	92.19%	100.00%

It was assumed that corn/soybeans could compete with other types of land use at any location except in the “exclusionary zones”. An exclusionary zone is an area that is not likely to change to agricultural use (e.g., urban, wildlife protection areas, and water bodies). It was also assumed that the urban area would stay static during the modeling time period since the urban area in North Dakota has not expanded notably during the past few decades (data source: <http://factfinder2.census.gov/>).

4.3. Model calibration and validation

The LTM was calibrated by training the ANN using the aggregated biofuel cropland maps for 1999/2000 and 2004/2005. The calibration generated multiple candidate ANNs with different ATF weights, as well as a set of simulated biofuel cropland change maps (during 1999/2000–2004/2005) created using these ANNs. The selection of an ANN for later simulation and forecast was based on the allocation disagreement and figure of merit that measure the goodness of fit between the observed and simulated changes. The ANN with the smallest allocation disagreement and largest figure of merit values was then selected as the calibrated model. The simulated biofuel cropland change map for 1999/2000–2004/2005 generated by the calibrated model is shown

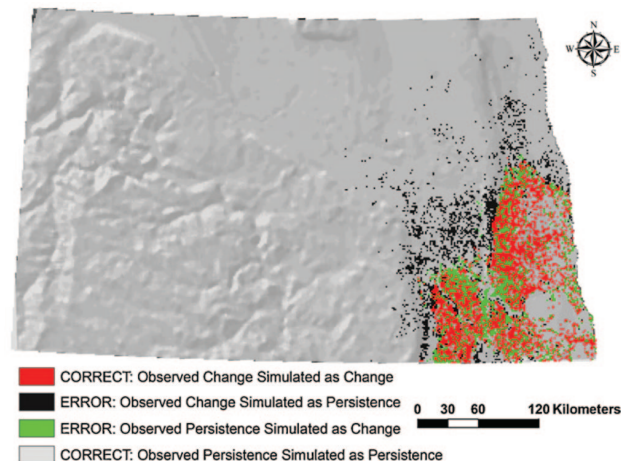
**Figure 8.** Correctness and error map for the calibration based on the 1999/2000 cropland reference map, the 2004/2005 cropland reference map, and the 2004/2005 simulated cropland change map.

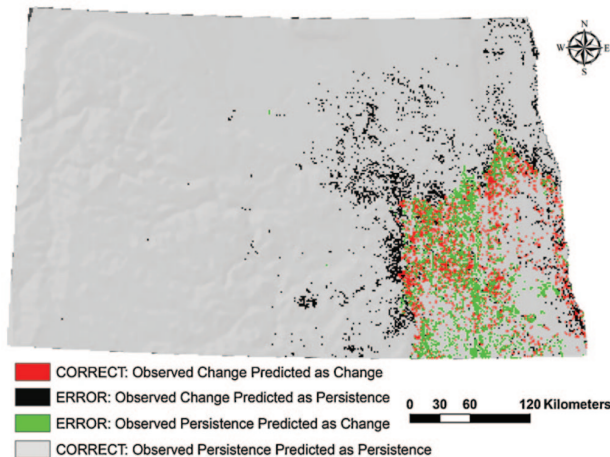
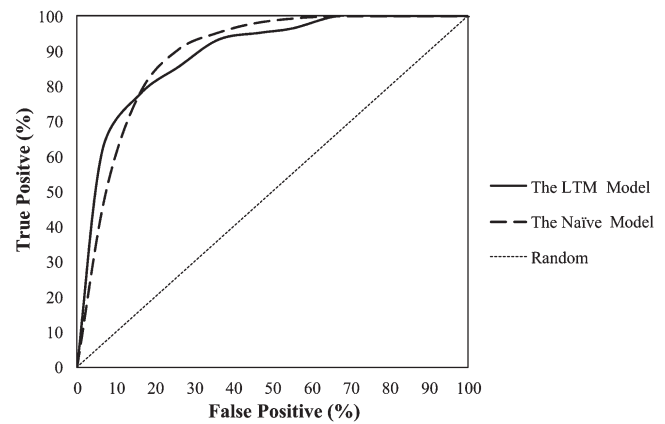
Table 4. Two-by-two contingency table showing the proportion of grid cells in a map of reality versus a map of validation scenario (2004/2005–2010/2011).

Model	Reality		
	Change	Persistence	Total
Change	2.38%	3.15%	5.53%
Persistence	3.15%	91.32%	94.47%
Total	5.53%	94.47%	100.00%

in Figure 8, along with the reference maps (i.e., aggregated historical data) for 1999/2000 and 2004/2005 (a three-map comparison). Derived from Figure 8, Table 3 show the percent areas of observed change simulated correctly, observed persistence simulated correctly, observed change simulated as persistence, and observed persistence simulated as persistence, illustrating the allocation disagreement between the simulated and actual cropland changes during the calibration period. As mentioned before, the quantity disagreement was eliminated because the number of cells to convert in the calibration process was set to be the same as that of the historical data. The resultant allocation disagreement was 5.56% and figure of merit was 47.5%. Table 3 also shows that 5.03% of all pixels were correctly simulated to change during the calibration period, 2.78% were observed persistence but simulated as change, and 2.78% were observed change but simulated as persistence, indicating there is more error ($2.78\% + 2.78\% = 5.56\%$) than the correctly simulated change (5.03%) at the pixel level, which is not unusual for LULCC modeling (Pontius et al., 2008).

To validate the model and its calibration, the cropland change between 2004/2005 and 2010/2011 was predicted using the calibrated model (Figure 9). Table 4 shows the allocation disagreement between the predicted and actual cropland changes during 2004/2005–2010/2011. Again, the quantity disagreement was none as the number of cells to convert in the validation process was set to be the same as the observed changes. The allocation disagreement was 6.3% and figure of merit 27.4%. Again, there is more error than the correctly simulated change at the pixel level.

To establish a fair baseline regarding the modeling performance (Pontius and Millones, 2011), a multi-factor index model was implemented as a naïve model, which linearly combines the same six driving factors (each factor was normalized to a 0–1 scale) as the suitability and chooses the highest suitable cells to convert until the number of simulated changed cells reaches the number of observed changed cells. The cropland changes for the validation period (i.e., 2004/2005–2010/2011) were predicted using the naïve model, and were then compared with the historical

**Figure 9.** Correctness and error map for the validation based on the 2004/2005 cropland reference map, the 2010/2011 cropland reference map, and the 2010/2011 predicted cropland change map.**Figure 10.** ROC curves to validate potential additional land changes between 2004/2005 and 2010/2011 for three suitability maps based on (1) the LTM model, (2) the naïve model, and (3) random location.

maps. The resulting allocation disagreement and figure of merit were 9.12% and 19.51%, respectively, indicating its performance inferior to the LTM-based model. This is because the LTM's ANN performed better than the naïve model's linear combination in capturing the nonlinear complex relationships between the driving factors and cropland change.

The relative operating characteristic (ROC) method was also applied to the validation. The ROC is a statistic used to measure the goodness-of-fit of the validation simulation (Pontius and Schneider, 2011; Pontius and Batchu, 2003). It was computed by generating multiple cropland change maps using different threshold values of suitability (which eventually lead to different numbers of cells to convert) and comparing the simulated maps with a Boolean 2010/2011 cropland change map (with 2004/2005 croplands excluded). The resultant ROC curves are shown in Figure 10. The areas under the curves (AUCs) for the calibrated LTM-based model and naïve model were quite similar. This similarity can be explained by the fact that both the LTM-based model and the naïve model portrayed most of the North Dakota study area (i.e., areas not in the southeast part of the state) as having low suitability for biofuels crop cultivation.

5. Future scenario projection

After the validation process, the calibrated model was then used to forecast the future corn/soybeans croplands for the year 2020. First, two scenarios of possible increases in corn/soybeans croplands were generated using simple extrapolation based on the historical agricultural statistics (National Agricultural Statistics Service, 2010). Linear and Weibull equation formulas were used to fit the historical corn/soybeans acreage, respectively. These formulas were assumed to project potential increases of corn and soybeans acreage in 2020 for fast and slow expansion scenarios, respectively (Figure 11). The fast scenario assumes that corn/soybeans will continue to increase rapidly due to high biofuel demands in the near future. The slow scenarios assume the rate of increase will gradually decrease. This may reflect the growing competition between biofuels and food industries for corn and soybeans (Horelik, 2008). Although other growth models with higher growth rates than the linear model may be potentially useful, a linear model still provided a realistic high expansion scenario, especially with increasing land competition among crops. The results of fast and slow projections were used to calculate the numbers of cells to be converted to corn/soybeans croplands during the period between 2010/2011 and 2020 under these scenarios. The LTM was

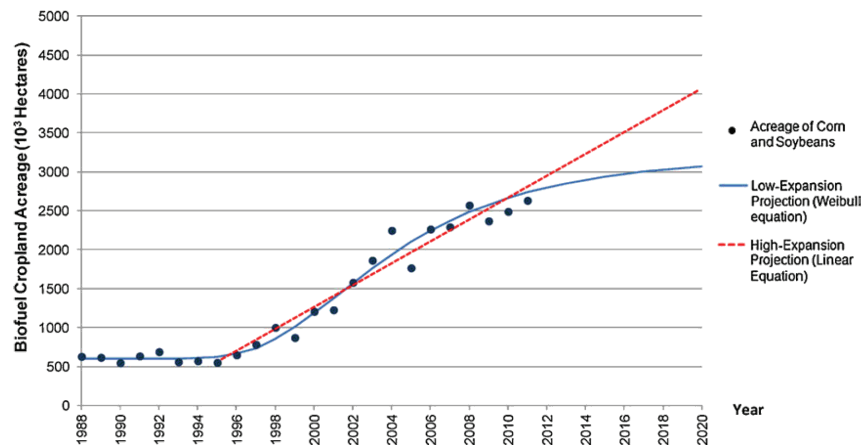


Figure 11. Two scenarios of cropland change.

then activated to spatially distribute the cells using the 2010/2011 cropland map as an initialization map and assuming that the cells with higher conversion probabilities will convert first.

The projections of future corn/soybeans changes are shown in Figure 12. As expected, under both scenarios, most changes in corn/soybeans croplands were projected to occur in eastern North Dakota, the Lake Agassiz Plain and the Northern Glaciated Plains, where the soil is generally fertile, the topography is relatively flat, and the climate is warmer and wetter than in the western parts of the state.

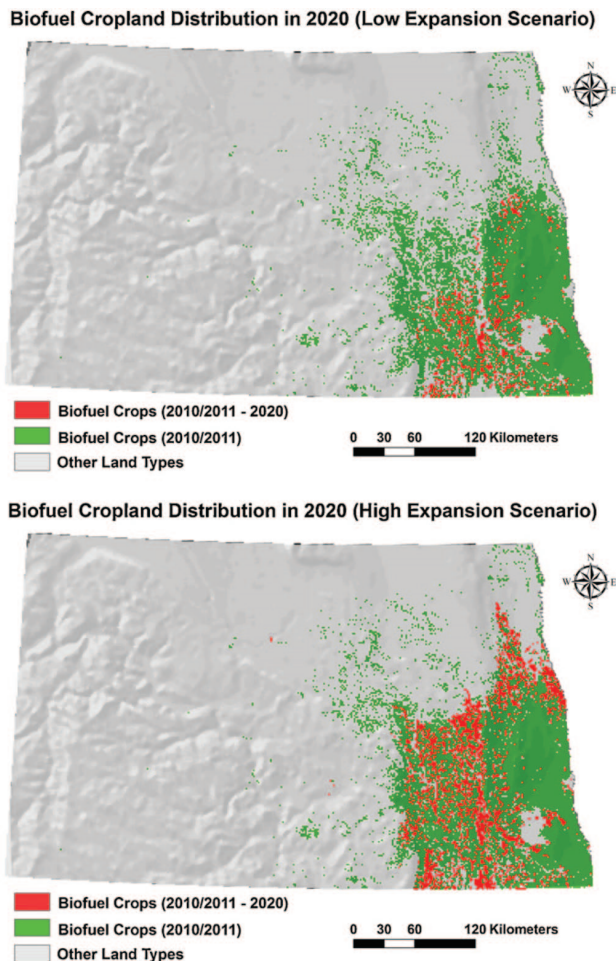


Figure 12. Modeled corn/soybeans cropland changes in 2020 under low and high expansion scenarios.

6. Discussion

6.1. Understanding the modeling results

Two major trends can be identified in the modeling results: (1) biofuel cropland changes were more concentrated in southeastern North Dakota in the simulation map than in the actual historical map; and (2) biofuel croplands appear to be expanding north-westward from southeastern North Dakota. Based on the driving factors used in this model, southeastern North Dakota is the most suitable area for agricultural cultivation because of its fertile soils, low and flat topography, and warm and wet climate. As the demand for biofuel crops keeps increasing, biofuel croplands expand from highly suitable areas (i.e., southeastern North Dakota) to moderately suitable areas (i.e., central North Dakota). The expansion of biofuel croplands follows the gradient of suitability for biofuel crop cultivation.

6.2. Evaluation of the modeling approach

The LTM was used in this study because it has been proven to be capable of dealing with complex relationships and noisy datasets. The LTM was implemented using the procedures similar to those that have been used in urban land-use change modeling (Pijanowski et al., 2002, 2005; Tang et al., 2005), the major difference being in the data preprocessing (i.e., spatial and temporal aggregation) to smooth the variations caused by crop rotations and other short-term, local fluctuations, such as farmers' planting decisions.

The LTM-based modeling methodology tested in this study is more spatially explicit than other approaches based on highly aggregated units such as countries or statistical districts (de la Torre Ugarte and Ray, 2000; Smeets et al., 2006). The modeling results can be easily used in other models such as, for example, groundwater pollution risk models, in order to assess the environmental impacts of biofuel cropland expansion. Through the spatial and temporal aggregations, our modeling framework is capable of distinguishing long-term regional trends in LULCC from frequent local changes (e.g., crop rotations) in cropping practices. Such short-term changes have often been overlooked in many other studies, since they used either spatially broad units or very large cell size (e.g., 10 km × 10 km cells used by Scheffran and BenDor, 2009). Hallmann and Verburg (2011), for example, acknowledged the importance of crop rotation but did not provide a viable solution to such a problem in their research.

However, this modeling approach was observed to have a number of limitations. For example, the ANN embedded in the

LTM is essentially a “black-box” and, therefore, it is difficult to identify and quantify the causal relationships between driving factors and LULCC. In addition, the LTM does not account for the annual LULCC dynamics that may occur during the simulation period. For example, the spatial distribution of land use in a specific year may affect the land use in the following year, and the driving factors may change every year as well.

Several assumptions were made to simplify the modeling process. First, biofuel crops were assumed to expand from areas with high agricultural suitability to ones with lower suitability. As mentioned earlier, the model essentially generates a suitability map for croplands, and then selects the cells exhibiting the highest suitability to convert. Thus, non-biofuel crops in areas of high suitability may be among first to be replaced by biofuel crops. The model goodness-of-fit indicates this assumption is valid. Second, the climatic factors used in the model were long-term (1970–2000) mean values. Climate change may affect the patterns of biofuel crops and other land uses. Changes in temperature and precipitation in the future may make some areas more or less suitable for cultivating biofuel crops, and hence affect the projection of future biofuel crops. Finally, the relationships between the driving factors and land use change were assumed to be static over time. Once the ANN is trained using historical data (i.e., 1999/2000–2004/2005), the weights of ATFs between nodes do not change during the simulation/validation (i.e., 1999/2000–2010/2011) and forecasting (i.e., 2010/2011–2020) periods, which means the ANN’s functional relationships between the input layer and output layer stay static.

The simulations suggest that biofuels-related LULCC is most likely to occur in eastern North Dakota. This is consistent with the trends observed in recent decades and described in other studies (see, for example, Wallander et al., 2011; Galle et al., 2009). The approach used in this study could be adapted to simulate other types of LULCC and could be applied in other regions that are likely to experience large agricultural LULCC.

7. Future research

The current implementation and tests suggest several avenues for future research, including the following:

1. To make the modeling location-specific, a threshold of 40% was used to define biofuel cropland cells in North Dakota. As a test, the model was recalibrated using different thresholds. The figure of merit of the recalibrated model was found improved as the threshold decreased (as shown in Figure 13). This is consistent with the discussion in Section 3, which is that applying a lower threshold produces more biofuel cropland cells and hence provides a larger land-use change contrast for the model, but exaggerates the quantity of corn/soybeans croplands. A threshold of 40% appears to provide a fair fit, although additional research should be undertaken to choose the most appropriate threshold.
2. The demand for biofuels was not computed using state-of-the-art socio-economic models. As mentioned above, the quantity module is flexible enough to link to more sophisticated external models. Thus, future work should focus on improving such forecasts by incorporating cost and risk factors to the crop production using aggregated projection models such as IMAGE/GTAP (Hallmann and Verburg, 2011) and POLYSYS (de la Torre Ugarte and Ray, 2000).
3. The premise underlying the modeling factor selection in the case study is that topography, soil and climate conditions govern the agricultural suitability for corn/soybeans, and corn/

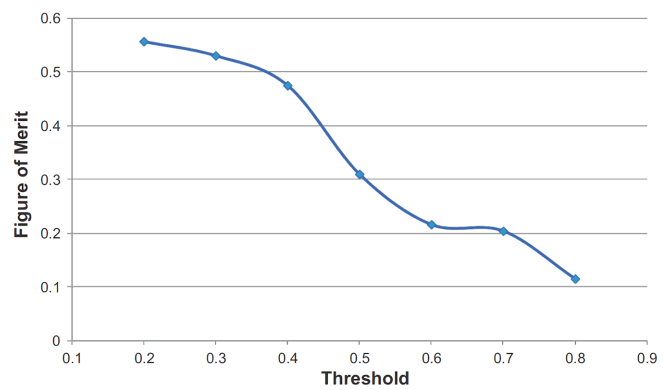


Figure 13. Figure of merit versus threshold for defining a biofuel cropland cell.

soybeans will most likely expand from areas with higher suitability to those with lower suitability. These factors were selected based on a review of related literature (e.g., Bowen and Hollinger, 2002; Kravchenko and Bullock, 2000). At the scale of a state, these factors were assumed to drive the long-term, regional LULCC for corn/soybeans. However, there are clearly other factors contributing to the expansion of biofuel croplands. Future research is needed to identify and utilize such factors in the modeling.

4. The current modeling framework is based on one-way relationships between the driving factors and LULCC, and ignores the feedback impacts of LULCC on the driving factors, for example, converting various types of lands into uniform biofuel croplands may change the local-scale climate. Lacking of spatio-temporal interactions between the driving factors and LULCC is a common problem in many current LULCC models. Future research should focus on the development of modeling mechanisms to incorporate such interactions.
5. Competing use of corn and soybeans as bioenergy and food crops was not distinguished in this study because our focus was the modeling of overall expansion trend. The land occupation of corn and soybeans for sole biofuel production will be separated in a future research. Also the interaction among specific crops such as corn, wheat and canola will be explicitly modeled in an improved modeling framework.

8. Conclusions

In this research, a framework for modeling and forecasting biofuels-related LULCC was developed and evaluated. The combination of a quantity module and an ANN-based spatial allocation module provides a means to simulate the amount and locations of biofuel cropland changes. The modeling framework was implemented to simulate the recent history of corn/soybeans cropland expansion in North Dakota, and forecast the future scenarios of biofuel cropland distribution in year 2020.

When compared with other biofuel cropland models, the approach proposed in this study was found to have two major advantages: (1) the approach is location-specific and, thus, produces output that can be exported directly to other models that demand such data (e.g., water quality models), and (2) through the spatial and temporal aggregations, the model distinguishes long-term, regional trends in LULCC from frequent short-term, local changes (e.g., crop rotations) in cropping practices. Nevertheless, several issues remain to be resolved: (1) the modeling errors, when compared to correctly predicted changes, are still too large, (2) the ANN’s black-box structure hinders further analy-

sis of the factors driving LULCC, and (3) the annual dynamics of LULCC are not explicitly modeled. Further research needs to be conducted to address these issues.

Models such as that proposed in this study can provide natural resources decision makers a means to understand the geographic extent of future cropland change in order to better address accompanying environmental consequences. As the demand for biofuels continues to grow, more land is likely to be converted to biofuel crops. This model, if coupled with environmental impact models, could assist decision makers in formulating land use policies and developing environmental management strategies to address impacts of biofuel cropland expansion.

Acknowledgments — We gratefully acknowledge the support for this research provided by the School of Natural Resources and the Center for Advanced Land Management Information Technologies, University of Nebraska-Lincoln. We also would like to thank the anonymous reviewers for their valuable comments and suggestions to improve this paper.

References

- Bian, L., 1997. Multiscale nature of spatial data in scaling up environmental models. In: Quattrochi, D.A., Goodchild, M.F. (Eds.), *Scale in Remote Sensing and GIS*. CRC Lewis, Boca Raton, FL, pp. 3–26.
- Bowen, C.R., Hollinger, S.E., 2002. Alternative Crops Web Site. Illinois State Water Survey, Champaign, IL, Retrieved January 13, 2010 from <http://www.sws.uiuc.edu/data/altcrops/>
- Brooke, R., Fogel, G., Glaser, A., Griffin, E., Johnson, K., 2009. Corn Ethanol and Wildlife –How Increases in Corn Plantings are Affecting Habitat and Wildlife in the Prairie Pothole Region. A University of Michigan Study Published by the Natural Wildlife Federation, Ann Arbor, MI.
- Cao, C., Lam, N.S., 1997. Understanding the scale and resolution effects in remote sensing and GIS. In: Quattrochi, D.A., Goodchild, M.F. (Eds.), *Scale in Remote Sensing and GIS*. CRC Lewis, Boca Raton, FL, pp. 57–72.
- Carriquiry, M., 2007. U.S. biodiesel production: recent developments and prospects. *Iowa Ag Rev.* 13 (2), 8, 9, 11.
- Clarke, K.C., Hoppen, S., Gaydos, L., 1997. A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay Area. *Environ. Plan. B* 24, 247–261.
- de la Torre Ugarte, D.G., Ray, D.E., 2000. Biomass and bioenergy applications of the POLYSYS modeling framework. *Biomass Bioenergy* 18 (4), 291–308.
- de Oliveira, M.E.D., Vaughan, B.E., Rykiel Jr., E.J., 2005. Ethanol as fuel: energy, carbon dioxide balances, and ecological footprint. *BioScience* 55, 593–603.
- Economic Research Service (U.S. Department of Agriculture), 2009. North Dakota Fact Sheet, Retrieved June 23, 2009 from <http://www.ers.usda.gov/Data/FarmIncome/FinfdmXls.htm>
- Evans, J.M., Fletcher, R.J., Alavalapati, J., 2010. Using species distribution models to identify suitable areas for biofuel feedstock production. *GCB Bioenergy* 2 (2), 63–78.
- Fan, F., Wang, Y., Wang, Z., 2007. Temporal and spatial change detecting (1998–2003) and predicting of land use and land cover in core corridor of Pearl River Delta (China) by using TM and ETM+ images. *Environ. Monit. Assess.* 137 (1–3), 127–147.
- Galle, A.M., Linz, G.M., Homan, H.J., Bleier, W.J., 2009. Avian use of harvested crop fields in North Dakota during spring migration. *West. N. Am. Naturalist* 69, 491–500.
- Goodchild, M.F., Dale, A.Q., 1997. Scale, multiscaling, remote sensing, and GIS. In: Quattrochi, D.A., Goodchild, M.F. (Eds.), *Scale in Remote Sensing and GIS*. CRC Lewis, Boca Raton, FL, pp. 1–11.
- Gosav, S., Praisler, M., 2008. Artificial Neural Networks Built for the Recognition of Illicit Amphetamines using a Concatenated Database, vol. XXXI. The Annals of the Dunărea de Jos University of Galați, Fascicle II (Mathematics, Physics, Chemistry and Informatics), pp. 150–155.
- Griffin, T., 2004. Know your Soil–Soil Organic Matter. University of Maine Cooperative Extension Publication, Retrieved September 29, 2010 from <http://umaine.edu/publications/2288e/>
- Haupt, S.E., Pasini, A., Marzban, C., 2009. Artificial Intelligence Methods in the Environmental Sciences. Springer, New York, NY.
- Hallmann, F., Verburg, P.H., 2011. Spatially explicit modelling of biofuel crops in Europe. *Biomass Bioenergy* 35 (6), 2411–2424.
- Hilbert, D.W., Ostendorf, B., 2001. The utility of artificial neural networks for modelling the distribution of vegetation in past, present and future climates. *Ecol. Model.* 146 (1–3), 311–327.
- Horelik, E.N., 2008. Biofuels in the United States transportation sector: public policy and its effects on the industry. *J. Eng. Public Policy* 9; online at <http://www.wisemintern.org>
- Kajita, Y., Toi, S., Tatsumi, H., 2005. Prediction of land use change in urbanization control districts using neural network—a case study of regional hub city in Japan. In: *ERSA Conference Papers ersa05p415*. European Regional Science Association, Vienna, Austria.
- Kennedy, D., 2007. The biofuels conundrum. *Science* 316 (5824), 515.
- Kravchenko, A.N., Bullock, D.G., 2000. Correlation of corn and soybean grain yield with topography and soil properties. *Agron. J.* 92, 75–83.
- Lambin, E.F., Rounsevell, M.D.A., Geist, H.J., 2000. Are agricultural land-use models able to predict changes in land-use intensity? *Agric. Ecosyst. Environ.* 82 (1), 321–331.
- Mas, J.F., Puig, H., Palacio, J.L., Sosa-Lopez, A., 2004. Modelling deforestation using GIS and artificial neural networks. *Environ. Model. Softw.* 19, 461–471.
- National Agricultural Statistics Service, U.S. Department of Agriculture, 2010. Acreage, Online, retrieved August 5, 2010 from: <http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1000>
- Pijanowski, B.C., Brown, D.G., Shellito, B.A., Manik, G.A., 2002. Using neural networks and GIS to forecast land use changes: a land transformation model. *Comput. Environ. Urban Syst.* 26 (6), 553–575.
- Pijanowski, B., Pithadia, S., Alexandridis, K., Shellito, B., 2005. Calibrating a neural network-based urban change model for two metropolitan areas of the Upper Midwest of the United States. *Int. J. Geogr. Inform. Sci.* 19 (2), 197–215.
- Pontius, R., Millones, M., 2011. Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *Int. J. Remote Sens.* 32 (15), 4407–4429.
- Pontius, R., Schneider, L., 2011. Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. *Agric. Ecosyst. Environ.* 85, 239–248.
- Pontius, R.G., Boersma, W., Castella, J., Clarke, K., Nijs, T., Dietzel, C., Duan, Z., Fotsing, E., Goldstein, N., Kok, K., Koomen, E., Lippitt, C.D., McConnell, W., Mohd Sood, A., Pijanowski, B., Pithadia, S., Sweeney, S., Trung, T.N., Veldkamp, A.T., Verburg, P.H., 2008. Comparing the input, output, and validation maps for several models of land change. *Ann. Regional Sci.* 42 (1), 11–37.
- Pontius, R., Batchu, K., 2003. Using the relative operating characteristic to quantify certainty in prediction of location of land cover change in India. *Trans. GIS* 7 (4), 467–484.
- Schnitkey, G., 2010. Changes in Crop Acres Since Freedom to Farm. The Farm Economics Facts and Opinions (FEFOs), 10–05.
- Secchi, S., Babcock, B.A., 2007. Impact of high crop prices on environmental quality: A case of Iowa and the Conservation Reserve Program. *Iowa Ag Rev.* 13 (2), 1–3.
- Scheffran, J., BenDor, T., 2009. Bioenergy and land use: a spatial-agent dynamic model of energy crop production in Illinois. *Int. J. Environ. Pollut.* 39 (1/2), 4–27.
- Smeets, E.M.W., Faaij, A.P.C., Lewandowski, I.M., Turkenburg, W.C., 2006. A bottom-up assessment and review of global bio-energy potentials to 2050. *Prog. Energy Combust. Sci.* 33 (1), 56–106.
- Tang, Z., Engel, B.A., Pijanowski, B.C., Lim, K.J., 2005. Forecasting land use change and its environmental impact at a watershed scale. *J. Environ. Manage.* 76 (1), 35–45.
- Thomas, M.A., Engel, B.A., Chaubey, I., 2009. Water quality impacts of corn production to meet biofuel demands. *J. Environ. Eng.* 135, 1123–1135.
- Tuck, G., Glendinning, M.J., Smith, P., House, J.I., Wattenbach, M., 2006. The potential distribution of bioenergy crops in Europe under present and future climate. *Biomass Bioenergy* 30 (3), 183–197.
- Turner, M.G., 1990. Spatial and temporal analysis of landscape patterns. *Landscape Ecol.* 4 (1), 21–30.
- U.S. Global Change Research Program, 2000. U.S. National Assessment of the Potential Consequences of Climate Variability and Change: A Detailed Overview of the Consequences of Climate Change and Mechanisms for Adaptation; <http://www.usgcrp.gov/usgcrp/nacc/> (accessed 2002)
- Verburg, P.H., de Koning, G.H.J., Kok, K., Veldkamp, A., Bouma, J., 1999. A spatial explicit allocation procedure for modelling the pattern of land use change based upon actual land use. *Ecol. Model.* 116 (1), 45–61.
- Verburg, A., Soepboer, P., Veldkamp, W., Limpiada, R., Espaldon, V., Mastura, S.S.A., 2002. Modeling the spatial dynamics of regional land use: the CLUE-S model. *Environ. Manage.* 30 (3), 391–405.
- Woodard, G., 2009. Alt-fuel development may impact natural resources. *Southwest Hydrol.* 8 (2), 8–9.
- Wallander, S., Roger, C., Cynthia, N., 2011. The Ethanol Decade: An Expansion of U.S. Corn Production, 2000–09, Economic Information Bulletin 117982. U.S. Department of Agriculture, Economic Research Service.